DOCUMENT RESUME

ED 247 239 TM 840 384

AUTHOR Mullis, Ina V. S.
TITLE Issues and Answers: What Will NAEP Tell Us?
PUB DATE Apr 84
NOTE 17p.; Paper presented at the Joint Annual Meetings of the American Educational Research Association and the National Council on Measurement in Education (New Orleans, LA, April 23-27, 1984).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Descriptive (141)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Achievement Tests; *Educational Assessment; Educational Objectives; Educational Policy; Elementary Secondary Education; Intellectual Disciplines; National Programs; Student Characteristics; Test Construction; Test Results; *Test Use
IDENTIFIERS *National Assessment of Educational Progress

ABSTRACT

The National Assessment of Educational Progress (NAEP) approach to developing assessments is described. NAEP will solicit views concerning assessment objectives and use the consensus approach. NAEP results can impact educational policy and school and classroom practices. Several policies have been adopted to provide improved information about specific subject area achievement. Subject areas will be selected to address issues of national concern and information needs. Each biennial assessment will include four subject areas. Subject area assessments will be more closely integrated and related and will include reading. Innovative item development will be assured through adoption of several policies. Each subject area assessment will contain open ended items and assessment of higher order reasoning skills will be emphasized. Assessment results will be more useful to a variety of audiences as a result of NAEP adoption of new policies. An increased scope of student level and school level background information will be collected. The collection of background information about teachers that can be directly related to student achievement has been initiated. Types of information may be classified as program related variables, impact of policies and practices, effect of teaching strategies and school curricular policies related to specific subject areas, and equal learning opportunity. (DWH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

I. V. S. Mullis

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED247239

ISSUES AND ANSWERS: WHAT WILL NAEP TELL US?*

Ina V. S. Mullis
National Assessment of Educational Progress
Educational Testing Service

What should be assessed? How should it be assessed? As a "national report card" on educational progress, the assessment should gather information that most Americans consider important. The skills, attitudes and understandings assessed should be basic to schooling and measured in ways that testing professionals and the public alike consider sound. Further, the results must be valid and reliable, both from professional and lay viewpoints. Having met criteria for relevance and validity, the assessment results and instruments must also prove useful to a wide range of people dedicated to strengthening our educational system.

With these considerations in mind, the NAEP approach to developing assessments rests on a number of principles and assumptions. Primary among these are the following:

1) NAEP is a public enterprise. It should solicit views from people in all walks of life about what should be assessed; the materials and processes should be open to public scrutiny; and it should rely on a consensus approach to develop objectives and goals.

2) NAEP can impact educational policy. It has an obligation to address issues being raised by major national studies urging educational improvement.

*Paper presented at the joint annual meeting of the American Educational Research Association and National Council for Measurement in Education, New Orleans, 1984.

TM 840 384

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

3) NAEP can impact school and classroom practice. It can address issues related to school effectiveness and improvement and suggest changes or actions that might be implemented by teachers and school administrators.

4) NAEP methodology should meet the highest technical standards. It must continue to pioneer the development of new kinds of items and capitalize on the benefits of state-of-the art psychometric theory and procedures.

5) NAEP must learn efficient ways to measure educational progress. It must be attentive not only to providing more information and increasing the utility of that information, but also to conserving resources.

To enhance various subject area findings, be more helpful to educational practitioners and address policy issues that transcend particular subject areas, NAEP has an improved new design. The key features of this design include sampling grade as well as age, BIB spiralling and application of item response theory. In addition, NAEP has taken innovative steps toward improving the utility of information through the kinds of measures that will be included in assessments and the relationships among those measures. These innovations are explained in the following sections.

NAEP Plans for Selecting Subject Areas

To provide improved information about achievement in specific subject areas and the relationships in learning and achievement among subject areas, NAEP has adopted the following policies:

- Subject areas will be selected to address issues of national concerns and information needs.
- At least four subject areas will be included in each biennial assessment.
- Assessments of various subject areas will be more closely integrated and related.
- Reading will be included in each biennial assessment.

Although the selection of subject areas for the current 1983-84 assessment was completed prior to the NAEP transition to ETS, it should be noted that this assessment of reading and writing involves a natural grouping of two compatible subject areas that previously had never been assessed together. The way ETS has redesigned the assessment, spiralling items across booklets and IRT scaling, will make it possible for us to investigate relationships in learning and achievement among these important communications skills.

Communications skills are of current national concern from a human resource perspective. "A Nation at Risk" states, "The teaching of English should equip graduates to: a) comprehend what they read, b) write well-organized, effective papers, and c) listen effectively and discuss ideas intelligently."

Although NAEP will not be able to address listening skills, through a variety of reading comprehension items, a variety of open-ended writing tasks and tasks that require both reading and writing, NAEP will be able to answer questions about levels of comprehension, effectiveness of written communication and how well students can discuss their ideas in writing. The 1983-84 writing assessment includes more writing tasks than in the past--15

exercises at each age/grade level--covering a variety of informative, persuasive and literary writing purposes and situations. In addition, as part of the reading assessment students are asked to interpret, analyze and evaluate passages in an open-ended format.

We are currently engaged in a massive open-ended scoring effort of over 200,000 responses that are BIB spiralled through the 1985-86 booklets and include responses from reassessed items administered in previous assessments. These results, along with the information from the objective reading exercises, should permit us to report levels of performance, how levels of performance have changed over time, and for this assessment, the relationships among particular reading and writing skills and abilities.

NAEP will assess Mathematics, Reading, Science and Computer Competence in the 1985-86 assessment. We refer to this assessment as the "Technology Package," since it includes two conceptual areas underlying most technologies and one important technological application. As an integrated assessment "package" mathematics, science and computers complement each other very well, and this assessment will enable NAEP to investigate the relationships in learning and achievement in these subject areas.

In addition, information about these subjects is extremely relevant to policy concerns, both separately and in concert. Many major national reports about the quality of education in our country have raised questions about the level of science and mathematics achievement in our schools and our nation's ability to meet the requirements of a technological society.

Mathematics is a national concern, both as a basic functional skill and as an area necessary to our country's economic growth. For example, between 1975 and 1980, public four-year colleges were forced to increase their remedial course offerings by 72 percent and many school districts are in the

process of increasing mathematics graduation requirements. Similarly, there is widespread concern that we, as a country, are not well prepared in the area of science. Data collected by the National Academy of Science indicate that science study in Japan, Russia, China and West Germany begins in the sixth grade. Their report notes, "The time spent on science, based on class hours, is approximately three times that spent by even the most science oriented students in the United States." Computer competency is extremely policy relevant because of the economic advantages for people with these skills, the burgeoning presence of computing and computer courses in schools, and the concerns of business and industry. Further, the 15th Annual Gallup Poll of the Public's Attitudes Toward the Public Schools shows that 80 percent of the parents with children in schools that do not have computers say they would like computers available for their children.

Given the current national concerns about economic leadership in tomorrow's world, the 1985-86 combination of assessments has the potential for enormously valuable analysis and reporting. Not only do science and computer competence complement mathematics and reading to create a unified assessment, but each is also individually powerful and gives NAEP the opportunity to be innovative and engaged in a "cutting-edge" effort to provide the most enlightening and useful information possible.

Since reading performance, more than any other area of achievement, is considered an indicator of student success and readiness for more advanced study in other fields, reading will be included in each biennial assessment to provide an important barometer of national educational progress. Frequent reading assessments will benefit both the analysis and reporting of NAEP data by increasing the number of data points and helping NAEP establish a firm trend in performance. NAEP has failed to hold the public's attention in the

past, partially because assessments, in any given subject matter, have been too infrequent and sporadic. Reporting performance results every two years in at least one subject area will give the public some regularly recurring information.

In addition, with the new spiralling design, student reading performance can be linked to student performance in all the subject areas in each assessment. This link between reading and other subject areas in each assessment also will enable NAEP to relate performance in different subject areas assessed in different years.

Finally, assessing reading will permit a routine analysis of alternating cohort samples. The 1983-84, 1987-88, and 1991-92 samples would represent one set of birth-year cohorts, while the 1985-86, 1989-90, and 1993-94 samples would represent an alternate wave of student cohorts. With reading common to all waves, cohort differences can be appraised and calibrated.

In summary, it should be noted that NAEP plans for policy responsive selection of integrated assessments including four or more subject areas will, of course, continue to fulfill the 1978 legislation authorizing NAEP which mandates that NAEP "collect and report at least once every five years data assessing the performance of students at various age or grade levels in each of the areas of reading, writing, and mathematics." As explained previously, plans call for reading to be assessed every two years. Also plans are to assess mathematics and writing every four years. Thus, the assessment schedule yields the following picture:

1983-84	Reading	Writing		
1985-86	Reading	Mathematics	Science	Computer
1987-88	Reading	Writing	?	?
1989-90	Reading	Mathematics	?	?

The above strategy appears to create a number of attractive alternatives for future assessments. Many integrated assessments about relevant concerns can be developed around the core of reading and writing or reading and mathematics.

NAEP Plans for Innovative Exercise (Item) Development

To insure that NAEP will continue to provide the information most relevant to educational policy-makers and practitioners, NAEP has adopted the following policies regarding exercise development:

- Each subject-area assessment will continue to contain open-ended items.
- Emphasis will be placed on assessing higher-order reasoning skills.
- Research will be conducted regarding the efficiency of measures, both in terms of the potential for administation techniques and the economies of measurements that eventually will be gained from IRT scaling.

The temptation to save money by developing assessments consisting only of multiple-choice items will be staunchly resisted in light of problems with face validity and an inability to assess many complex skills.

As part of the development effort for the 1985-86 assessment, staff is implementing a planned, focused effort to assess higher-order reasoning competencies, both within and across subject areas. NAEP's ability to provide such information will address issues of major concern to educators, business leaders, and parents.

Some higher-order skills and abilities may be subject-area specific, such as understanding the relationships underlying the key concepts within each specialized discipline; while others, such as pattern recognition, formulating hypotheses, applying concepts and generalizations to new situations and evaluating information, are skills that cut across subject areas in

interesting and meaningful ways. For example, Academic Preparation for College: What Students Need to Know and Be Able to Do, published by the College Board, states that students should be able to "draw reasonable conclusions from information found in various sources, whether written, spoken, tabular, or graphic." This could include tasks such as interpreting written works (reading), interpreting data from maps (social studies), generating hypotheses from experimental data (science) and interpreting data from tables and graphs (mathematics).

The NAEP tasks will relate to knowledge and concepts relevant to subject areas. A "content free" thinking assessment containing mental games or gymnastics is not envisioned. Rather we hope either to provide students with the knowledge and information necessary to perform the given task or to construct situations whereby we can determine whether lack of success probably resulted from not having the information or from not having an appropriate strategy to solve the problem.

NAEP recognizes that such an endeavor is very difficult and must be planned and approached carefully. However, NAEP is equally certain that exercises should require students to actually demonstrate their thinking skills. For too long, testing in general has relied on asking students to react, rather than act. Of course, this does not mean that NAEP assessments will not continue to include the more traditional exercises measuring knowledge and attitudes. It is that priority in new development efforts will be given to higher-order reasoning exercises, particularly those that measure skills that apply across subject areas.

Work towards this goal began early this year by convening a combined meeting of the four Learning Area Committees (Reading, Mathematics, Science and Computer Competence). At a joint session members of all four groups were

given the same charge—to focus on higher-order skills in guiding the design of each particular assessment and then to work together to investigate the commonalities in assessments that might permit developing exercises across areas.

There was agreement about the goal, and the groups worked independently and effectively, basing their definitions of subject-matter domains on awareness of current research and their own experience. Each developed, along with subject-matter content definitions, a dimension responsive to assessing higher-order tasks. Although, these are in draft stages and subject to further refinement, the mathematics group suggested assessing the cognitive levels of understanding and comprehension, routine applications and problem solving/reasoning. The science team developed a cube using three broad terms—content, context and cognitive levels—with the various cognitive levels being knowledge, ability to use knowledge and ability to integrate, synthesize and apply knowledge. The reading experts suggested the identification of information, integration/use of information, extension of information and making critical judgments. Finally, the computer committee worked in two areas of major concern, applications and computer science. The cognitive levels currently proposed for applications are knowledge, operation and use; while those for computer science are knowledge, skills, understanding, routine application and problem/solving reasoning.

As noted, these initial definitions are currently undergoing a stringent review process and the challenge of representing these objectives with valid exercises still lies ahead. Meanwhile, NAEP is searching the literature, consulting with experts and exploring a variety of administration techniques including individual or small-group assessments involving "hands-on" exercises. In addition, we are investigating ways to capitalize on NAEP's use

of trained exercise administrators and to implement more innovative methods of open-ended assessment based on NAEP's vast experience and history in this area.

Finally, NAEP must engage in a planned, focused research effort over the next assessments that will enable us to determine which assessment methods will most effectively and efficiently address various cognitive levels. It may be that a change in emphasis is needed in that NAEP has not utilized resources effectively and has used expensive methods to collect information about achievement related to recognition and recall or even some abilities related to higher-order skills that could be more efficiently and just as effectively measured through carefully designed objective measures. Meanwhile, we have neglected a concerted effort to evaluate the questions we ask students in terms of the nature of the cognitive demands required. NAEP is committed to giving this effort in hopes of providing improved information about students' thinking abilities. We know it will not be easy and field tests may yield disappointing results. Yet, the hope is to advance both the art of measurement and the utility of NAEP information.

NAEP Plans for Background Information

To make NAEP more useful to a variety of audiences, assessment results must address issues central to improving the quality of education. In addition, it makes sense that the most useful information about these issues would go beyond describing or explaining differences in achievement, and suggest changes or actions that might be implemented. Thus, NAEP has adopted the following policies:

- NAEP has substantially increased the scope of student-level background information collected.

- NAEP has substantially increased the scope of school-level background information collected.
- NAEP has initiated the collection of background information about teachers that can be directly tied to student achievement.

The types of information collected may be classified into four broad areas: 1) program-related variables, 2) the impact of policies and practices suggested by the school effectiveness research, 3) the effect of teaching strategies and school curricular policies and environments related to specific subject areas, and 4) equality of opportunity to learn.

More specifically, utility of NAEP achievement data is being addressed through collected information about a variety of student, classroom/teacher and school variables related to:

1) the impact of participation in various programs on achievement (for example, Chapter I, bilingual education, programs for handicapped students, programs for handicapped students, vocational education and gifted and talented programs);
2) the impact of home on achievement (for example, reading materials in the home, reading and writing activity in the home, homework done, television viewing, computer access and use, working mother, family size and composition, language spoken in the home, level of parents' education, after-school supervision and free time activities);
3) the quality of the education work force (teacher certification, teacher training, role of the principal, size of staff, morale of staff, experience of staff and in-service training opportunities) and that relationship with achievement;

4) the effect of various school problems (absenteeism, discipline, students transfers, vandalism, lack of commitment and motivation and student use of drugs or alcohol) on achievement;

5) the time available for instruction (length of school year, day and class periods) and that relationship to achievement;

6) the relationship between ethnic composition of schools and student achievement;

7) the relationship between student performance and minimal competency testing programs or graduation test requirements;

8) the effect of resources as evidenced by pupil-teacher ratios, class size, teacher class load, or various equipment and facilities on achievement;

9) how particular curricular approaches (materials, textbooks, delivery systems, homework assigned and evaluation techniques) relate to student performance in particular subjects;

10) the relationship between time spent in school on the subject (courses taken, class time spent and frequency of reading and writing activities) and student performance in the subject;

11) the effect of overall curriculum focus (courses taken, classes enrolled in and overall program) on achievement;

12) the relationship between preschool experience and student performance;

13) the relationship between students' perceptions and attitudes toward education, school and specific subject matter and performance;

14) how time spent outside of class on subject matter-related activities appears to relate to achievement;

15) how school characteristics may relate to student performance (region and size and type of community in which the school is located; the kind of school, e.g. public/private; range of grade levels included and the size of school and number of faculty); and

16) how student performance might be influenced by school policies regarding ability grouping, remediation, special services.

In summary, NAEP should be thought of not merely as a social indicator, but as a tool to identfy problems and suggest areas of research concerning educational progress. Timely analyses of achievement data in relation to relevant background variables should suggest provisional interpretations and promising leads that merit further investigation and research.

NAEP Plans for Additional Assessment Activities

In addition to improved information resulting from timely and integrated subject-area assessments, innovative exercise development focused on higher-order reasoning skills and the inclusion of an increased number of policy-relevant background variables within the context of the existing assessment framework, NAEP will seek to conduct special assessments or probes in the years between regularly scheduled assessments.

The first of these activities will be an assessment of young adults, ages 21-25, to provide a profile of the literacy skills of these adults including both high school dropouts and college graduates. This additional grant from NIE marks a return of NAEP to an original goal of the program, assessing young adults.

This group is particularly critical since most are about to assume responsibility for the nation's work. As recent products of our education and training programs, they are still young enough to maximize the benefits of corrective remedial efforts.

The project will build upon recent efforts to address adult literacy issues by reviewing current work and holding two conferences to help reach a consensus on the definition of "literacy," set objectives and select sample exercises. As with the in-school assessments, particular care will also be given to identifying the background characteristics and educational experiences associated with various levels of literacy.

The data generated should provide information to policymakers and program operators concerned with providing young adults with the reading, writing, computational, problem-solving and work awareness skills needed by individuals and society for social and employment purposes.

NAEP's Process for Developing Assessments

It should be stressed that NAEP in its effort to address new issues and provide more useful information will rely on an assessment development process that will be, if anything, even more rigorous than in the past. In addition to careful attention to the use of numerous external consultants representing various diverse perspectives, backgrounds and constituencies, we will use empirical techniques to determine the effectiveness of each item.

The legislation that created NAEP is clear in its statement that objectives should be arrived at through a consensus process, and our intention is to follow that specification scrupulously. The procedures followed for determining objectives for assessments are essentially those followed by NAEP in the past. Objectives from previous assessments were reviewed by a wide range of educators and specialists in the field selected to represent differing points of view, geographical locations, backgrounds and educational environments. With the guidance of the Learning Area Committees, selected again with great care, the results of reviews were synthesized and objectives were modified and updated. This new edition of the objectives will in turn be reviewed by practitioners and members of the lay public from around the country for their reactions and opinions. Several redrafting and review phases will be conducted, with the Learning Area Committees participating in each phase of the process.

Similarly, a carefully tested series of steps essentially those followed by NAEP in the past, will be used to develop exercises. Based on the revised objectives, existing assessment items, previous assessment data and past experience, NAEP staff and external consultants will develop exercise prototypes and specifications. Exercises will, for most part, be developed during item writing workshops and conferences conducted by staff and involving external consultants. The items that have been created will be reviewed and revised by staff and by external reviewers considered subject-matter and measurement experts. In addition, all exercises will be reviewed for bias. Further language editing and sensitivity reviews will be conducted following ETS quality control procedures and all materials will be cleared by NIE, FEDAC and OMB. National field tests will be conducted and the results scored and analyzed. Based on the results of field tests the entire process of

modifying, reviewing and editing exercises will be repeated. With the help of staff and outside reviewers, the Learning Area Committees will select the specific exercises that will be included in the assessment.

Summary

NAEP will now assess at least four subject areas every other year. By selecting relevant and related subject areas, this will cost less and give NAEP the power to examine relationships across subject areas. As the NAEP legislation mandates, assessment objectives will be developed through a consensus process that meets the needs of the public and respects NAEP's public-funding status. The new assessments will present a series of exercises that can be analyzed to measure of number of learning factors in various combinations and examine the interrelationships among questions. Development efforts will focus on innovative open-ended assessment of reasoning skills. However, to remain relevant and efficient, NAEP must find the resources for expanded assessment opportunities through continual redesign and implementation of more cost-effective procedures.

Increased attention to the collection of student, school and classroom/teacher variables from students, school officials and teachers will enable NAEP to address current national concerns that focus on school effectiveness questions, program policies, teaching strategies and equity issues. All development procedures will be conducted to meet high technical standards.